Evolutionary inevitability of sexual antagonism

Tim Connallon and Andrew G. Clark

Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA

Sexual antagonism, whereby mutations are favourable in one sex and disadvantageous in the other, is common in natural populations, yet the root causes of sexual antagonism are rarely considered in evolutionary theories of adaptation. Here, we explore the evolutionary consequences of sex-differential selection and genotype-by-sex interactions for adaptation in species with separate sexes. We show that sexual antagonism emerges naturally from sex differences in the direction of selection on phenotypes expressed by both sexes or from sex-by-genotype interactions affecting the expression of such phenotypes. Moreover, modest sex differences in selection or genotype-by-sex effects profoundly influence the long-term evolutionary trajectories of populations with separate sexes, as these conditions trigger the evolution of strong sexual antagonism as a by-product of adaptively driven evolutionary change. The theory demonstrates that sexual antagonism is an inescapable by-product of adaptation in species with separate sexes, whether or not selection favours evolutionary divergence between males and females.

1. Introduction

Species with separate sexes face two important evolutionary challenges that can limit their abilities to adapt to a changing environment. First, discrete sexes are common among complex organisms, in which pleiotropic effects can constrain adaptation [1–3]. Second, patterns of selection and the phenotypic effects of mutations each differ between the sexes [4–6], which can decouple the genetic basis of male versus female fitness. When male and female fitness is decoupled, mutations benefitting one sex will sometimes be deleterious to the other (i.e. their fitness effects are ‘sexually antagonistic’ [7]), which diminishes a population’s ability to respond to selection through fixation of beneficial mutations.

An emerging body of data implies that sexual antagonism is an important feature of animal and plant populations [8–11]. Nevertheless, several fundamental questions regarding sexual antagonism remain unaddressed. For example, how often does sexual antagonism arise owing to sex differences in the direction of selection? How severely does sexual antagonism limit the rate of adaptation in each sex, and how do such constraints evolve over time? How does environmental change mediate opportunities for sexual antagonism, including differences between poorly versus well-adapted populations?

These questions can be addressed with theoretical models that effectively bridge the divide between empirically measurable properties of mutation and selection, and the evolutionary genetic patterns and processes that we seek to understand. Quantitative genetics theory has proved extremely useful for quantifying short-term constraints to sex-specific adaptive evolution [12], yet we currently lack clear theoretical predictions about the underlying population genetics of sexual antagonism, including the fraction of mutations that has sexually antagonistic fitness effects, and the role of sexual antagonism in shaping the long-term evolutionary trajectories of males and females in adaptively evolving populations. With these issues in mind, we developed a two-sex extension of Fisher’s geometric model [13], and used it to characterize the sex-specific distribution of mutant fitness effects, and the population genetic dynamics of adaptation in species with separate sexes.
Fisher’s geometric model in species with separate sexes

Fisher’s original model provides a simplified mapping between genotype, phenotype and fitness that captures the basic biological details of adaptation within a complex organism: (i) mutations cause random changes within multi-dimensional trait space; (ii) selection favours those mutations that move the system of traits closer to an optimal phenotypic value that maximizes fitness; and (iii) a population approaches its fitness optimum by successively fixing beneficial mutations [14,15]. Although alternative modes of evolutionary change—e.g. polygenic adaptation involving modest allele frequency shifts at many loci—may also contribute to adaptation over short evolutionary time-scales, such short-term responses may be weak when pleiotropic constraints—including sexual antagonism—are severe [16–18]. In such cases, adaptation is likely to be dominated by the fixation of beneficial mutations, provided directional selection is sustained over time [19,20]. Fisher’s model also makes simplifying assumptions about the mutational architecture of traits, yet its original assumptions can be relaxed without greatly altering the model’s basic predictions [21–23] and these predictions are adept at explaining a variety of interesting empirical patterns that emerge from the study of real organisms (e.g. [24–28]).

In extending Fisher’s model to a two-sex system, we characterize the evolution of n traits that are each expressed, but not necessarily equally, in males and females (see the electronic supplementary material). Sets of trait values are depicted using Cartesian coordinates, with each set of coordinates representing an individual’s genetically determined position in n-dimensional phenotypic space. Following prior work, we assume that fitness depends on the Euclidean distance of each sex to its optimum, and mutational effects are unbiased in direction (uniformly oriented in n-dimensional space [13,29,30]). These assumptions are not restrictive as long as we interpret n as the ‘effective complexity’ of a species (i.e. the effective number of traits can be less than the actual number [21–23]). To simplify the presentation, we model mutation and evolution in a haploid population; our results also apply to diploids, with the caveat that mutations contributing to adaptation in diploids will sometimes involve a transient, balanced polymorphic state [30]. Finally, although we focus on evolution in dioecious (gonochoristic) species with distinct sexes, sexual antagonism may also manifest within hermaphrodites, by way of allocation trade-offs between male and female reproductive structures in simultaneous hermaphrodites, or antagonistic pleiotropy between male and female stages in sequential hermaphrodites [31]. Our results may imperfectly characterize adaptation in hermaphrodites, yet our underlying mathematical framework is sufficiently flexible to permit future theoretical extensions to mating systems that we do not specifically consider here.

Patterns of selection depend on the distance and orientation of each sex to its phenotypic optimum. Let \( z_m \) and \( z_f \) represent the Euclidean distance between the optimal phenotype and the actual phenotype expressed by each sex (subscripts, m and f, hereafter refer to ‘male’ and ‘female’). Fitness functions for each sex are Gaussian, with \( w_m = \exp(-\omega_m z_m^2) \) and \( w_f = \exp(-\omega_f z_f^2) \), and \( \omega_m \) and \( \omega_f \) are positive constants specifying the strength of the fitness decline with increased distance from the optimum. The relative orientation of each sex to its optimum—the direction of selection—can be described by a pair of vectors that extend from the current locations of each sex within phenotypic space to the location of its optimum (figure 1). The strength of selection is defined as \( \beta_m = \frac{\partial \ln(w_m)}{\partial z_m} = 2\omega_m z_m \) and \( \beta_f = \frac{\partial \ln(w_f)}{\partial z_f} = 2\omega_f z_f \), which parallel the definition of the selection gradient from quantitative genetics (see [32, pp. 122–123]). \( \theta_{\text{mut}} \) represents the angle between male and female selection vectors, with \( \rho_{\text{mut}} = \cos(\theta_{\text{mut}}) \) representing the correlation between male and female orientations of directional selection (\( -1 < \cos(\theta_{\text{mut}}) < 1 \), with \( \rho_{\text{mut}} = 1 \) representing identical directions of selection; see [33]).

Phenotypic effects of mutations are similarly described using paired vectors, with each mutation having a unique magnitude (\( r_m \) and \( r_f \)) and orientation with each sex (figure 1). For each mutation, \( \theta_{\text{mut}} \) is a random variable, which represents the angle between male and female orientations of phenotypic change caused by the mutation. For mutations with specified magnitude \( r_m \) and \( r_f \), the phenotypic correlation between the sexes for each trait axis is given by \( \rho_{\text{mut}} \). In the two-sex extension of Fisher’s model, there is a geometric relationship between \( \rho_{\text{mut}} \) and the angle between mutation vectors: \( \rho_{\text{mut}} = \cos(\theta_{\text{mut}}) \), where the angle brackets denote the mean among random mutations (see the electronic supplementary material).

Unique mutations arise randomly within a population that is initially fixed at each locus for a resident (wild-type) allele. For a given mutation, its sex-specific fitness effect
in the \( j \)th sex is \( s_j = \frac{\log \left( \frac{w_j(z_j^+)}{w_j(z_j^-)} \right)}{\log(1+1)} \), where \( z_j^+ \) and \( z_j^- \) represent distances to the optimum for wild-type and mutant individuals, respectively. Letting \( t_j = \log(1+1) \) represent the logarithm of relative fitness, \( f(t_{m}, t_1; r_{m}, r_1) \) is the bivariate distribution of \( t_m \) and \( t_1 \) among mutations with arbitrary magnitude, \( r_m \) and \( r_1 \) for \( r_m(r_1 > 0) \). In many dimensions (\( n > 1 \)), this distribution converges to a bivariate normal distribution, with marginal mean and variance \( \left( t_j \right) = -aw_j^2 \) and \( \text{var}(t_j) = 2aw_j^2n \), respectively, and between-sex correlation coefficient of \( \rho_{md} = \cos(\theta_{md})/(\cos(\theta_{md})) = \rho_{md}\rho_{md} \) (see the electronic supplementary material), noting that \( s_{mv} s_l \) will also be bivariate normal in the limit: \( s_{mv} s_l \to 0 \). This simple identity for \( \rho_{md} \) implies that genotype-by-sex interactions and sex-differential selection contribute multiplicatively to the fitness effect correlation between the sexes. Moreover, either factor is sufficient to generate an imperfect genetic correlation between male and female fitness.

Adaptive evolutionary change involves the sequential fixation of mutations with sex-averaged beneficial fitness effects \( s_{avg} = s_m/2 + s_l/2 > 0 \). Similar to previous work in Fisher’s geometric model (e.g. [29]), we characterize evolutionary change in a large population where beneficial mutations are rare, selection coefficients of beneficial alleles are small \( (s_m, s_l \ll 1) \); see [34], and therefore each mutation’s probability of fixation depends only on its net fitness effect. Provided the sex-averaged fitness effect is positive \( s_{avg} > 0 \), a mutation’s probability of fixation is \( 1 - \exp(-2s_{avg}) \), and otherwise its fixation probability is zero. Mutations that are successfully fixed trigger a change in the location of males and females within phenotypic space. Following each substitution, new distances and relative orientations of males and females to their optima may be calculated (i.e. new values of \( s_{mv} z_m \theta_{md} \)), and the next step of adaptation can proceed from this new population state (see the electronic supplementary material).

Because substitutions in Fisher’s model are discrete (see above and [29]), each time point during evolution is associated with specific values for \( s_{mv} z_m \theta_{md} \); and these terms may therefore be treated as population parameters appropriate for a given time point. We assume throughout that the phenotypic distributions of mutations remain constant during evolution (distributions of \( r_{mv} r_1 \) and \( \theta_{md} \) remain constant). Therefore, \( \rho_{md} \) can be treated as a population parameter, albeit one that can evolve over the course of an adaptive walk. Accordingly, we first consider the sex-specific fitness effects of new mutations and the rate of adaptation, for an arbitrary population with parameters \( z_{mv} z_l \theta_{md} \). We then characterize the long-term evolution of such populations as they adapt, with each sex approaching its fitness optimum.

### 3. Results and discussion

Adaptation requires beneficial genetic variation, which serves as the fuel for evolutionary change. In populations with separate sexes, random mutations can be deleterious to both sexes, beneficial to both, or beneficial to one sex and deleterious to the other (‘sexually antagonistic’). Among mutations that are beneficial to at least one sex \( s_m > 0 \) and/or \( s_l > 0 \), the fraction that is sexually antagonistic is:

\[
\frac{f_{SA}}{f_{SA}} = \frac{\Pr(s_m < 0)}{1 - \Pr(s_m > 0; s_l < 0)}.
\]

which, in the limit of small mutation size, reduces to a simple function of the fitness effect correlation \( \rho_{md} \):

\[
\lim_{r_{mv}, r_{l} \to 0} f_{SA} = \frac{2\pi - 4\sin^{-1} (\rho_{md})}{3\pi - 2\sin^{-1} (\rho_{md})}.
\]

Equation (3.1b) approximates the minimum proportion of beneficial mutations that are sexually antagonistic (see the electronic supplementary material). Numerical evaluation of equation (3.1), along with exact computer simulations across a distribution of \( r_m \) and \( r_1 \) values, shows that \( f_{SA} \) increases with dimensionality and mutation size—factors that also increase the ‘scaled size’ of mutations in Fisher’s geometric model (i.e. the scaled mutation size in sex \( j \) is defined as \( x_j = r_j/(2z_j) \); see [13,29]). Overall, sexual antagonism is pervasive, and is a dominant feature of beneficial mutations, across the entire range of \( \rho_{md} \) (figure 2).

Adaptation represents a compromise between fitness benefits and costs to males and females, with selection favouring mutations with positive, sex-averaged fitness effects (i.e. \( s_{avg} = s_m/2 + s_l/2 > 0 \)). Rates of adaptation in each sex can be approximated by taking into account both the mutation rate to positively selected alleles and the fixation probabilities of individual mutations [1]. To quantify the effect of an imperfect fitness correlation on the rate of adaptation, constraint is expressed as the ratio of the expected rate of adaptation in a population with imperfectly correlated male and female fitness effects \( \rho_{md} < 1 \) to the rate of adaptation in an ideal population with perfectly correlated fitness effects \( \rho_{md} = 1 \). We formally define this constraint as \( C_j = (dz_j/dt \mid \rho_{md} < 1)/(dz_j/dt \mid \rho_{md} = 1) \), where \( dz_j/dt \) is the rate of evolutionary change of sex \( j \) towards its optimum (see the electronic supplementary material; \( C_j = 1 \) in the absence of sexually antagonistic constraints, and otherwise, \( C_j < 1 \)). In the limit of small mutation size \( r_{mv}, r_{l} \to 0 \), these ratios in males and females are approximately:

\[
C_m = 1 - (1 - \rho_{md}) \frac{r_{m} \beta_{l}}{r_{m} \beta_{l} + r_{m} \beta_{m}}
\]

and

\[
C_l = 1 - (1 - \rho_{md}) \frac{r_{m} \beta_{m}}{r_{m} \beta_{m} + r_{l} \beta_{l}}
\]

which compares well with exact computer simulations of small to moderately sized mutations, and it underestimates the true magnitude of constraint, particularly when scaled mutation sizes are large (figure 2, where \( x_j = r_j/(2z_j) \) is the scaled size). An imperfect fitness correlation between the sexes significantly decreases the rate of adaptation \( C_m \) and \( C_l \) can be substantially lower than one), with the relative magnitude of constraint dependent upon mutation and selection asymmetries between the sexes. With equally strong directional selection in each sex and similar-sized mutations \( r_{mv} \beta_{m} = r_{l} \beta_{l} \), males and females experience equal degrees of constraint: \( C_m = C_l = 1 - (1 - \rho_{md})/2 \). Asymmetries between the sexes \( r_{mv} \beta_{m} \neq r_{l} \beta_{l} \) shift the burden of constraint to the sex that experiences weaker selection or mutation (figure 2).

The results above describe features of selection and adaptation at single time points during evolution, and consequently they treat fitness effect distributions and sex-specific rates of adaptation and constraint as static rather than evolutionarily labile population properties. However, each mutation that is fixed during adaptation will carry the population to a new location in phenotypic space and will alter the distance and orientation of each sex to its optimum. Fitness effect distributions and subsequent opportunities...
for adaptation will therefore change following each genetic substitution. To characterize such changes during the course of adaptive evolution, we analysed adaptive walks of populations towards stationary male and female fitness optima (see the electronic supplementary material; figure S3). We focused on the strength and orientation of sex-specific directional selection ($\beta_m$, $\beta_f$ and $\rho_{mf}$)—quantities that directly impact the distribution of mutant fitness effects.

Despite an enormous range of possible initial conditions and sex-specific mutation and selection parametrizations, two general patterns emerge from the analysis. First, as long as mutant phenotypic effects are positively but imperfectly correlated between the sexes, as appears likely [6,36,37] (see below), the genetically coupled evolution of male and female phenotypes will inevitably generate sex-specific selection in opposing directions in phenotypic space (figure 3; i.e. adaptation eventually causes $\rho_{mf} < 0$, and thus, $\rho_{ad} < 0$ and $f_{SA} > 2/3$; see the electronic supplementary material). In other words, adaptation generates opposing selection on traits expressed by males and females, and widespread sexual antagonism among mutations that are individually beneficial to males or females. This outcome includes cases where directional selection is initially identical between the sexes (figure 3 and electronic supplementary material, figure S2). Second, initial differences in the strength of directional selection decrease during adaptation ($\beta_m/\beta_f \rightarrow 1$), which causes the severity of sex-specific adaptive constraints to converge over time (figure 3 and electronic supplementary material, figure S3). Given that $\rho_{ad}$ eventually evolves to be negative, sexual antagonism is expected to ultimately reduce the subsequent rate of adaptation in each sex by at least twofold ($C_0, C_m < 1/2$), with a strong majority of beneficial mutations having sexually antagonistic fitness effects ($f_{SA} > 2/3$).

Our model yields two general insights into the evolution of species with separate sexes. First, conditions for sexual antagonism among beneficial mutations are extremely permissive and readily emerge when directional selection or the phenotypic effects of mutations differ between the sexes. Indeed, the correlation of mutant fitness effects is equally sensitive to both factors, as reflected in the identity: $\rho_{ad} = \rho_{mf} \cos(\theta_{mut}) \cos(\theta_{ad})$. Second, decoupling the male and female fitness effects of random mutations significantly reduces the rate of adaptation of both sexes, with sexual antagonism and the magnitude of adaptive constraints increasing during the process of adaptation. Evolution of sexual antagonism occurs whether or not selection favours phenotypic divergence between the sexes (figure 3 and electronic supplementary material, figure S2). Opposing selection between the sexes—for homologous traits expressed by both sexes and mutations affecting these traits—is therefore an inescapable by-product of adaptation in populations with separate sexes.

The model is parametrized using measurable properties of mutation and selection and should therefore be useful for inferring the extent of sexual antagonism in populations conducive to such measures. Although sex-specific phenotypic properties of spontaneous mutations have yet to receive much attention, current data on genetic correlations between the sexes (based on quantitative genetic data [36,37], and mutation-accumulation experiments in Drosophila [6]) suggest that mutations are likely to have strong, positively correlated effects.
between the sexes. Estimates of selection on single traits suggest that opposing directional selection is common, and often of similar magnitude between the sexes [5,38]. Of the few studies estimating the angle between male and female directional selection on multiple traits, two support the general prediction that opposing selection readily evolves \((\rho_{\text{sel}} \approx -0.61 \text{ and } -0.73)\) in a moth and fly population, respectively [39,40], whereas a third reveals a positive correlation between male and female directional selection, albeit in a modern human population that has had little time to adapt to its current environment \((\rho_{\text{sel}} \approx 0.22 [41])\). To systematically test whether opposing selection between the sexes evolves during adaptation, future studies could contrast patterns of sex-specific directional selection in poorly versus well-adapted populations, with the latter expected to exhibit more strongly opposing selection between males and females (for a similar prediction, see [42]).

We have identified general conditions yielding sexually antagonistic selection, and although such antagonism does not require that selection favour evolutionary divergence between the sexes, the pervasiveness and intensity of

![Diagram](https://example.com/diagram.png)

**Figure 3.** Adaptation generates opposing directional selection between the sexes. (a) The average change in the orientation of directional selection in males and females during adaptation: \((\Delta \rho_{\text{sel}})\). Each data point represents the average change for 10 000 simulated substitutions, with simulation parameters \(z_m = z_f = 1/2, \omega_m = \omega_f = 1\), and mutation sizes drawn from a bivariate exponential distribution with marginal mean of \((r_m) = (r_f) = 0.1\) and between-sex correlation of \(\text{corr}(r_m, r_f) = \langle \cos(\theta_{\text{mut}}) \rangle = \rho_{\text{mut}}\). (b) The strength of directional selection in males and females \((\beta_m, \beta_f)\) and the relative orientation of directional selection between the sexes \((\theta_{\text{avg}})\) during adaptive walks. Each step of adaptation coincides with fixation of a positively selected mutation \((\text{with } s_{\text{avg}} > 0)\), and the solid lines show averages over 100 simulated adaptive walks \((\text{shaded area shows } \pm \text{ one standard deviation for } \rho_{\text{mut}})\). Two extreme initial conditions of the population are shown: \(\text{(i)}\) directional selection is initially identical in both sexes and phenotypic divergence between the sexes is not favoured \((\text{initial conditions: } z_m = z_f = 1/2, \theta_{\text{avg}} = 0, \rho_{\text{sel}} = 1)\); and \(\text{(ii)}\) selection favours phenotypic divergence between males and females \((\text{males are initially displaced from their optimum but females are not; initial conditions: } z_m = 1/2, z_f = 0)\). Mutation sizes were drawn from a bivariate exponential distribution with marginal means \(\langle r_m \rangle = \langle r_f \rangle = 0.05\), and \(\text{corr}(r_m, r_f) = \langle \cos(\theta_{\text{mut}}) \rangle = \rho_{\text{mut}} = 0.75\). Additional parameters include \(\omega_m = \omega_f = 1\). Additional results, and details of the simulation procedure, can be found in the electronic supplementary material.
antagonism is expected to increase when selection does favour sexual dimorphism. Discordance between the fitness landscapes of males and females—the root cause of selection for dimorphism—emerges from the unique interactions between each sex and its environment [8–11], which can lead to sex-differential opportunities for niche partitioning, reproductive investment, competition for mates or their susceptibilities to predation and parasitism. We emphasize the population genetic consequences of distinct male and female fitness landscapes, while placing no special emphasis on the specific ecological causes for divergent male and female optima. The ubiquity of sexually dimorphic phenotypes [43], including conspicuous differences between closely related species [44], attests to the biological fact of fitness landscape dimorphism and implies a central role for antagonistic selection during the process of adaptation within lineages and speciation between them.

Acknowledgements. We thank Deborah Charlesworth, Angela Early, Ted Morrow, Chris Steha, Akane Uesugi, Rob Unkless, Cris van Hout, and members of the Clark laboratory for valuable discussion and comments on an early version of the manuscript, Allen Orr for valuable correspondence during the initial stages of the project, and two anonymous referees for thoughtful comments and suggestions that substantially improved the paper.

Funding statement. This work was funded by NIH grant GM64590 to A.G.C. and A. B. Carvalho.

References


Supplementary Materials:

General model

Initial conditions, sex-specific optima, and the direction of selection. Suppose that males and females occupy arbitrary initial positions in n-dimensional space, \( A_m = (A_{m1}, A_{m2}, \ldots, A_{mn}) \) and \( A_f = (A_{f1}, A_{f2}, \ldots, A_{fn}) \), with each point represented by a vector of length \( n \) (i.e. a set of Cartesian coordinates). Points in space representing male and female optima are given by \( O_m = (O_{m1}, O_{m2}, \ldots, O_{mn}) \) and \( O_f = (O_{f1}, O_{f2}, \ldots, O_{fn}) \), and the vector pointing from \( A_j \) to \( O_j \) is of length \( z_{j,m} = z_{j,f} = \{m, f\}, \) i.e. the Euclidean distance between point \( A_j \) and \( O_j \). We are particularly interested in the angle between the male and female vectors, as this defines the orientation of directional selection in males relative to females. To calculate this angle, we can rescale the four points relative to the origin, such that: \( A_j' = (0, 0, \ldots, 0) \), \( O_{m,j}' = O_m - A_m \), and \( O_{f,j}' = O_f - A_f \). Rescaling yields a triangle with vertices \( A_j' \), \( O_{m,j}' \), and \( O_{f,j}' \), and edges of length \( z_m, z_f \), and \( d(O_{m,j}', O_{f,j}') \). From the law of cosines, the angle at vertex \( A_j' \) is given by:

\[
\theta_{sel} = \arccos \left( \frac{z_m^2 + z_f^2 - d^2(O_{m,j}', O_{f,j}')}{2z_m z_f} \right) = \arccos \left( \frac{\sum_{i=1}^n (O_{mi} - A_{mi})(O_{fi} - A_{fi})}{z_m z_f} \right)
\]

where \( d^2(O_{m,j}', O_{f,j}') = \sum_{i=1}^n [(O_{mi} - A_{mi}) - (O_{fi} - A_{fi})]^2 \). For convenience, we use rescaled positions in phenotypic space in subsequent results.

Mutation. The effect of a mutation can similarly be described using vectors pointing out from the initial positions, \( A_j \), in phenotypic space. For a mutation of magnitude \( r_j \) (that is, carriers of the mutation express a phenotype with Euclidean distance \( r_j \) away from \( A_j \)), we can generate random orientational changes in trait space using the following algorithm for simulating random,
uniformly distributed points on the surface of an \( n \)-dimensional hypersphere with radius \( r_j \) (S1, S2). We randomly draw \( n \) times from a standard normal distribution, labeling these variables \( y_i; i = \{1, \ldots, n\} \). For a mutation with magnitude \( r \), the change along the \( i \)th axis is given by \( x_i = r y_i / Y \), where \( Y^2 = y_1^2 + y_2^2 + \ldots + y_n^2 \). Note that \( r = \sqrt{\langle \sum y_i^2 \rangle} \), as required.

Since we are concerned with evolution with two sexes, and mutations can have different phenotypic effects per sex, we must describe each mutation using a pair of vectors (one for males and the other for females). For a mutation with effect \( r_m \) in males and \( r_f \) in females, we can generate random sex-specific vector orientations by conducting \( n \) draws from a bivariate standard normal distribution and standardizing using a similar approach as for the above case.

For males, draw \( n \) independent, random variables from a standard normal distribution, and label these as \( m_i \), with \( i = \{1, \ldots, n\} \). The corresponding vector of changes in \( n \)-dimensional space is \( x_m = r_m m_i / M \), where \( M^2 = m_1^2 + m_2^2 + \ldots + m_n^2 \). The mutation’s effect on individual phenotypic axes corresponds to \( x_{mi} = r_m m_i / M \). Assignment of an orientation for females depends on the mutational correlation between the sexes. To account for this, we draw \( n \) times from a normal distribution, with values conditioned on the terms of \( m_i \) for males. Letting \( \rho_{mut} \) represent the between sex correlation for mutational effects within each axis, then \( f_i \) is also normally distributed: \( \sim N[\rho_{mut} m_i, (1 - \rho_{mut}^2)] \). The change in individual female axes is \( x_{fi} = r_f f_i / F \), where \( F^2 = f_1^2 + f_2^2 + \ldots + f_n^2 \).

In the limit of high dimensionality, terms of \( M \) and \( F \) converge to \( \sqrt{n} \), a useful approximation that we use extensively in subsequent analytical results. This approximation stems from the relation \( M = \sqrt{\langle \sum m_i^2 \rangle} = \sqrt{n \bar{X}} \), where \( \bar{X} = (\sum m_i^2) / n \) represents the sample mean of \( n \) squared independent standard normal variables. \( \bar{X} \) is itself a random variable with mean and variance \( \mathbb{E}(m_i^2) = 1 \), and \( \text{var}(m_i^2)/n = 2/n \), respectively. The latter converges to zero in the limit of
high dimensionality, and thus, $M$ and $F \rightarrow \sqrt{n}$, for large $n$. In practice, the approximation works well for $n$ on the order of 10 or greater. Note, however, that all simulations results are based on the exact results using $M$ and $F$.

Taking into account the original positions of each sex, their optima, and the effects of the mutation on each sex, we can now characterize the phenotypic position of individuals carrying a new mutation, with respect to the new distances to each optimum. This new distance, in males and females, respectively, will be:

$$z_{m}' = \sqrt{\sum_i (O_{mi}' - x_{mi})^2} \approx \frac{2r_m}{\sqrt{n}} \sum_i O_{mi} m_i$$

$$z_{f}' = \sqrt{\sum_i (O_{fi}' - x_{fi})^2} \approx \frac{2r_f}{\sqrt{n}} \sum_i O_{fi} f_i$$

with the approximation appropriate for large $n$. The angle between male and female mutation vectors ($\theta_{mut}$, in the main text) is a function of changes in each axis. From the law of cosines, we have the relationship:

$$\sum_i (x_{mi} - x_{fi})^2 = r_m^2 + r_f^2 - 2r_m r_f \cos(\theta_{mut})$$

which reduces to $\sum_i x_{mi} x_{fi} = r_m r_f \cos(\theta_{mut})$. Taking the expectation, and again using the large-$n$ approximation [$M$ and $F \rightarrow \sqrt{n}$], we obtain the identity: $E[\cos(\theta_{mut})] = E(m_f) = \rho_{mut}$.

**Sex-specific distributions of mutant fitness effects.** Under a Gaussian fitness function, fitness of each sex is a simple function of the distance from its optimum. The fitness of sex $j$ is given by $w_j(z_j) = \exp(-\omega_j z_j^2)$, where $\omega_j$ is a scaling parameter that defines the rate of fitness decline about the optimum. As described in the main text, the selection coefficient of a random
mutation is \( s_j = (w_j(z_{j*})/w_j(z_{j+}) - 1) \), where \( z_{j+} \) and \( z_{j*} \) represent distances to the optimum for wild type and mutant individuals. Substituting eq. (1) (and relabeling \( z_j = z_{j+} \)) leads to:

\[
s_m \approx \exp \left[ -\omega_m \left( r_m^2 - \frac{2r_m}{\sqrt{n}} \sum_i O_{mi} m_i \right) \right] - 1
\]

\[
s_f \approx \exp \left[ -\omega_f \left( r_f^2 - \frac{2r_f}{\sqrt{n}} \sum_i O_{fi} f_i \right) \right] - 1
\]

The terms in exponents, \( t_m = \ln(1 + s_m) \) and \( t_f = \ln(1 + s_f) \), are linear functions of standard normal variables, making the overall terms normally distributed. The distribution of selection coefficients will therefore follow a bivariate shifted lognormal distribution, with marginal moments and between-sex correlation a function of the random normal variables \((m_i, f_i)\) and parameters \( z_j, \theta_{sel}, E[\cos(\theta_{mut})], n, r_j \) and \( \omega_j \).

Given that \( E(m_i) = E(f_i) = 0 \), and \( \text{var}(m_i) = \text{var}(f_i) = 1 \), the distribution of \( t_m \) and \( t_f \) is bivariate normal with marginal moments \( E(t_j) = -\omega_j r_j^2 \) and \( \text{var}(t_j) = 4(z_j \omega_j r_j)^2/n \). The covariance between \( t_m \) and \( t_f \) is:

\[
\text{cov}(t_m, t_f) = \frac{4\omega_m \omega_f r_m r_f}{n} \sum_i O_{mi} O_{fi}' \text{cov}(m_i, f_i) = \frac{4\omega_m \omega_f z_m z_f r_m r_f}{n} E[\cos(\theta_{mut})] \cos(\theta_{sel})
\]

From eq. (3) we see the influence of two constraints on adaptation in populations with separate sexes: (1) the phenotypic correlation between sexes with respect to random mutations (given by \( E[\cos(\theta_{mut})] \); \(-1 < E[\cos(\theta_{mut})] < 1\)); and (2) the correlation of directional selection (given by \( \cos(\theta_{sel}) \); \(-1 < \cos(\theta_{sel}) < 1\)). Moreover, there is a critical angle between the vectors pointing to male and female optima, for which the fitness effect distribution of mutations is completely independent between the sexes: \( \cos(\theta_{sel}) = 0 \) when \( \theta_{sel} = \pi/2 \), or 90 degrees.

We can also calculate the correlation between selection coefficients:
\[
\text{corr}(s_m, s_f) = \frac{\exp[\text{cov}(t_m, t_f)] - 1}{\sqrt{\{\exp[\text{var}(t_m)] - 1\}\{\exp[\text{var}(t_f)] - 1\}}}
\]  

(4),

where \(|\text{corr}(s_m, s_f)| < |\text{corr}(t_m, t_f)|\) for all \(0 < |\text{corr}(t_m, t_f)| < 1\) (S3), which serves to dampen the between sex correlation for the selection coefficients.

### Probability of sexually antagonistic mutations

Beneficial mutations can be unconditionally favorable in both sexes, or beneficial in only one sex and deleterious in the other (i.e., “sexually antagonistic”). The probability of a mutation having a beneficial effect in either sex will of course be constrained by the size of the mutation relative to the current distance to the optimum \((r_j/z_j)\) and the complexity of the organism \((n)\).

Thus, the fraction of mutations that are unconditionally beneficial, or sexually antagonistic, will be highly context-specific (for example, both mutation types may be common when mutations have small standardized phenotypic effects: \(r_j\sqrt{(n)}/2z_j << 1\); and rare when mutations have large effects: \(r_j\sqrt{(n)}/2z_j > 1\) (12, 18)). A more useful statistic – one that is less sensitive to the scaling of mutation size – is the proportion of beneficial mutations that are sexually antagonistic \((f_{SA}\) in the main text). We can calculate this proportion from the joint probability distribution of \(t_m\) and \(t_f\), which is approximately bivariate normal with density \(f(t_m, t_f; r_m, r_f)\). The probability that a mutation is beneficial to both sexes, \(\Pr(t_m > 0; t_f > 0)\), is represented by the positive orthant of \(f(t_m, t_f; r_m, r_f)\), i.e.:

\[
\Pr(t_m > 0, t_f > 0) = \frac{1}{2\pi\sqrt{1 - \rho_{mf}^2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-\frac{t_1^2 + t_2^2 - 2\rho_{mf}t_1t_2}{2(1 - \rho_{mf}^2)}\right) dt_1 dt_2
\]
where $x_m = r_m \sqrt{n}/(2z_m)$, and $x_f = r_f \sqrt{n}/(2z_f)$. The probability that a mutation is beneficial in at least one sex will be:

$$1 - \Pr(t_m < 0, t_f < 0) = \Pr(t_m > 0) + \Pr(t_f > 0) - \Pr(t_m > 0, t_f > 0)$$

$$= \frac{1}{\sqrt{2\pi}} \int_{x_m}^{\infty} e^{-t_m^2} \, dt + \frac{1}{\sqrt{2\pi}} \int_{x_f}^{\infty} e^{-t_f^2} \, dt - \frac{1}{2\pi} \int_{x_m}^{\infty} \int_{x_f}^{\infty} \exp \left( \frac{-t_1^2 + t_2^2 - 2 \rho_{mf} t_1 t_2}{2(1 - \rho_{mf}^2)} \right) \, dt_1 \, dt_2$$

Finally, the probability that a mutation is sexually antagonistic is:

$$\Pr(t_m / t_f < 0) = \Pr(t_m > 0) + \Pr(t_f > 0) - 2 \Pr(t_m > 0, t_f > 0)$$

$$= \frac{1}{\sqrt{2\pi}} \int_{x_m}^{\infty} e^{-t_m^2} \, dt + \frac{1}{\sqrt{2\pi}} \int_{x_f}^{\infty} e^{-t_f^2} \, dt - \frac{1}{2\pi} \int_{x_m}^{\infty} \int_{x_f}^{\infty} \exp \left( \frac{-t_1^2 + t_2^2 - 2 \rho_{mf} t_1 t_2}{2(1 - \rho_{mf}^2)} \right) \, dt_1 \, dt_2$$

In the limit $x_m, x_f \to 0$, the above equations reduce to:

$$\Pr(t_m > 0, t_f > 0) = \frac{1}{4} + \frac{\arcsin(\rho_{mf})}{2\pi}$$

$$1 - \Pr(t_m < 0, t_f < 0) = \frac{3}{4} - \frac{\arcsin(\rho_{mf})}{2\pi}$$

$$\Pr(t_m / t_f < 0) = \frac{1}{2} - \frac{\arcsin(\rho_{mf})}{\pi}$$

(see (S4), pp. 262-263). The proportion of beneficial mutations that are sexually antagonistic is

$$f_{SA} = \Pr(t_m/t_f < 0)/[1 - \Pr(t_m < 0, t_f < 0)]$$

which in the weak mutation limit simplifies to:

$$f_{SA} = \frac{2\pi - 4 \arcsin(\rho_{mf})}{3\pi - 2 \arcsin(\rho_{mf})}$$

which is presented as eq. (1) of the main text.

**Rates of adaptation**
Here we provide an extension of Orr’s (1) “cost of complexity” model to describe the rate of adaptation in each sex. We present our results in terms of the rate of movement of an arbitrary sex $j$ toward its optimum, i.e.: $dz_j/dt$. Characterizing the change in fitness (as in (1)) requires only a trivial modification of our results, i.e.: $dw_j/dt = (dz/dt)(dw/dz)$, where $w_j(z_j) = \exp(-\omega_j z_j^2)$, and $dw_j/dz_j = -2\omega_j z_j \exp(-\omega_j z_j^2)$.

In the subsequent analysis, we assume that mutations that potentially contribute to divergence each have small selection coefficients ($s_j \approx t_j$, as generally expected for beneficial mutations (1, 18, S5)). With small selection coefficients, changes in the distance to the optimum may be approximated using the useful relationship: $\Delta z_j \approx -s_j/(2\omega_j z_j)$; $\Delta z_j < 0$ when the substitution brings sex $j$ closer to its optimum; $\Delta z_j > 0$ when the substitution carries sex $j$ away from its optimum. The fixation probability of a mutation is:

$$Pr(fixed | s_m, s_f) = 1 - \exp[-(s_m + s_f)] = \frac{t_m + t_f}{1 - \exp[-N(s_m + s_f)] \exp[-N(t_m + t_f)]}$$

where $N$ is the population size ($N = N_e$ is assumed for simplicity (S6)). Following Orr (1), and given the stated assumptions, the rate of movement of sex $j$ toward its optimum is given by:

$$dz_j/dt = -N \mu \cdot E[Pr(fixed | t_m, t_f)] \cdot E(\Delta z_j | fixed) \quad (5)$$

The individual terms are:

$$E[Pr(fixed | t_m, t_f)] = \iiint Pr(fixed | t_m, t_f)f(t_m, t_f; r_m, r_f)dt_mdtdf,$$

and

$$E(\Delta z_j | fixed) = - E(t_j | fixed)/(2\omega_j z_j)$$

where:
\[ E(t_j \mid \text{fixed}) = \int \int t_j \Pr(\text{fixed} \mid t_m, t_f) f(t_m, t_f; r_m, r_f) dt_m dt_f \]

Substituting these terms into eq. (5), we obtain:

\[ \frac{dz_j}{dt} = - \frac{N\mu}{2\omega_j z_j} \int t_j(t_m + t_f) f(t_m, t_f; r_m, r_f) dt_m dt_f \]

\[ \int \frac{1}{1 - \exp[-N(t_m + t_f)]} \]  \hspace{1cm} (6),

In the limit of small mutation size, \( f(t_m, t_f; r_m, r_f) \) is approximately bivariate normal with mean and variance terms \( E(t_j) = 0 \), \( \text{var}(t_j) = (2\omega_j z_j)^2/n \). Then, using a Taylor series, we can expand eq. (6), and solve to second order in \( t_m \) and \( t_f \):

\[ \frac{dz_j}{dt} = - \frac{N\mu}{2\omega_j z_j} \times \frac{\text{var}(t_j) + \text{cov}(t_m, t_f)}{2} \]

This last result can be used to obtain eq. (2) of the main text.

**Change in \( \cos(\theta_{\text{sel}}) \) following an adaptively fixed substitution**

To characterize the change in the angle between selection vectors (\( i.e. \), following each adaptive substitution), we start by rescaling the positions of each sex with respect to their optima. Given the symmetry inherent in the model (mutation orientations are unbiased in phenotypic space, and sex-specific fitness is based on distance but not orientation to the optimum), we can rescale the positions of each sex relative to their optima at any arbitrary point in time (\( i.e. \), after each step during adaptation). For a given angle between selection vectors (\( \theta_{\text{sel}} \)), and male and female distances to their optima of \( z_m \) and \( z_f \), respectively, we rescale the phenotypic positions of males and females, and of their optima, to: \( A_m = A_f = (0, 0, \ldots, 0) \), \( O_m = (z_m \cos(\theta_{\text{sel}}), z_m \sin(\theta_{\text{sel}}), 0, \ldots, 0) \), and \( O_f = (z_f, 0, \ldots, 0) \). A random mutation will carry males and females to new positions:
\[ A_m = \{x_{m1}, x_{m2}, \ldots, x_{mn}\} \] and \[ A_f = \{x_{f1}, x_{f2}, \ldots, x_{fn}\} \], respectively. If the mutation becomes fixed, the new angle between selection vectors will be:

\[
\theta_{sel}^\prime = \arccos \left( \frac{z_m z_f \cos(\theta_{sel}) - z_f x_{m1} - z_m [\cos(\theta_{sel}) x_{f1} + \sin(\theta_{sel}) x_{f2}] + \sum_{i=1}^{n} x_{mi} x_{fi}}{\sqrt{(z_m^2 + r_m^2 - 2z_m [x_{m1} \cos(\theta_{sel}) + x_{m2} \sin(\theta_{sel})]) (z_f^2 + r_f^2 - 2z_f x_{f1})}} \right)
\]

Using the large-\(n\) approximation (see above), this modifies to:

\[
\theta_{sel}^\prime = \arccos \left( \frac{z_m z_f \cos(\theta_{sel}) - \frac{z_f r_m m_1}{\sqrt{n}} - \frac{z_m r_f [\cos(\theta_{sel}) f_1 + \sin(\theta_{sel}) f_2]}{\sqrt{n}} + \sum_{i=1}^{n} \frac{r_m r_f f_i}{n}}{\sqrt{(z_m^2 + r_m^2 - 2z_m [m_1 \cos(\theta_{sel}) + m_2 \sin(\theta_{sel})]) (z_f^2 + r_f^2 - 2z_f r_f f_1)}} \right)
\]

We also obtain the following relationships (each becomes useful below):

\[
\frac{t_m + \omega_m r_m^2}{2 \omega_m z_m r_m} = \frac{[m_1 \cos(\theta_{sel}) + m_2 \sin(\theta_{sel})]}{\sqrt{n}}
\]

and

\[
\frac{t_f + \omega_f r_f^2}{2 \omega_f z_f r_f} = \frac{f_1}{\sqrt{n}}
\]

We want to know the change in the angle between selection vectors following the next adaptively fixed substitution. Although we mostly resort to simulations to demonstrate changes to the angle (using the exact eq. (7)), analytical results can be reached for some specific cases. Two cases of particular interest are presented here: (1) the expected change to the angle when mutational effects are strongly and positively correlated between the sexes; and (2) the expected change when the selection direction is strongly and positively correlated between the sexes.

**Sex-differential selection with a strong phenotypic correlation of mutational effects.**

Consider the case where the direction of selection differs between males and females (\(\theta_{sel} > 0\),
and phenotypic correlations are strong (i.e., for the limiting cases: $\theta_{\text{mut}} \to 1; r_m \to r_f$). Thus, $m_i = f_i$, and $\rho_{mf} = \cos(\theta_{\text{sel}})$. The correlation between selection coefficients in the next generation is:

$$
\cos(\theta_{\text{sel}'}) = \frac{z_m z_f \cos(\theta_{\text{sel}}) - \frac{t_f}{2\omega_f} - \frac{t_m}{2\omega_m}}{\sqrt{\left(\frac{z_m}{z_f} - \frac{t_m}{\omega_m}\right)\left(\frac{z_f}{z_m} - \frac{t_f}{\omega_f}\right)}} = \cos(\theta_{\text{sel}}) - t_m \frac{\left(z_m - z_f \cos(\theta_{\text{sel}})\right)}{2\omega_m z_m z_f} - t_f \frac{\left(z_f - z_m \cos(\theta_{\text{sel}})\right)}{2\omega_f z_m z_f}
$$

with the approximation to first order in $t_m$ and $t_f$ (both assumed to be small). The angle increases when $\cos(\theta_{\text{sel}'}) < \cos(\theta_{\text{sel}})$. In the limit of small mutation size, we can substitute approximations for $E(t_m | \text{fixed})$ and $E(t_f | \text{fixed})$, in which case, the angle is expected to increase after the next substitution ($\cos(\theta_{\text{sel}})$ will decrease) when the following criterion is met: $(\omega_f z_f^2 + \omega_m z_m^2) \sin^2(\theta_{\text{sel}}) > 0$. This criterion is met as long as $0 < \theta_{\text{sel}} < \pi$, i.e.: $-1 < \cos(\theta_{\text{sel}}) < 1$. Thus, as long as the directional selection orientations differ slightly between the sexes ($\theta_{\text{sel}} > 0$), $\cos(\theta_{\text{sel}})$ is always expected to decrease and ultimately approach the boundary: $\cos(\theta_{\text{sel}}) = -1$.

**Sex-concordant selection and imperfect correlation of mutational effects.** Letting $\theta_{\text{sel}} = 0$, for $t_m, t_f$ small and $n$ large, we get the approximation:

$$
\cos(\theta_{\text{sel}'}) = 1 + \frac{\rho_{\text{mut}} r_m r_f z_f^2}{z_m z_f \sqrt{\left(1 - \frac{t_m}{z_m \omega_m}\right)\left(1 - \frac{t_f}{z_f \omega_f}\right)}}
$$

The angle increases when $\cos(\theta_{\text{sel}'}) < 1$, or $0 < (z_m r_m)^2 + (z_f r_f)^2 - 2 z_m z_f r_m r_f \rho_{\text{mut}}$, which is true for all $\rho_{\text{mut}} < 1$ (i.e., by analogy to the law of cosines). Thus, even when a population starts out with identical directional selection in males and females, an imperfect correlation of mutational effects on their phenotypes will guarantee that directional selection will differ between the sexes after the first fixed substitution.
Simulation Procedures

**Random mutations.** Each mutation was generated using a two-step process. First, for each mutation, the effect size in each sex ($r_m$ and $r_f$) was obtained by sampling from a bivariate gamma distribution with correlation coefficient of $\text{corr}(r_m, r_f)$ and equal marginal distributions in the two sexes, *i.e.* $r_j$ is from a gamma distribution with shape and scale parameters, $k$ and $\theta$, where $E(r_j) = k\theta$ and $\text{var}(r_j) = k\theta^2$. Values of $r_m$ and $r_f$ were obtained using the “mixture approach” algorithm, described in (S7). Second, mutational orientations were generated using the algorithm described above (S1, S2). Given the mutation’s size and orientation in phenotypic space, it is then straightforward to calculate its fitness effect in each sex, which is a function of the distance of mutant males and females from their respective optima.

**Fixed Substitutions.** Adaptively fixed substitutions were generated by filtering random mutations based on their relative probabilities of fixation. For each mutation arising uniquely in a population that is otherwise fixed for wild-type alleles, the probability of fixation for arbitrary strength of selection (and assuming that $N|s_m + s_f| >> 1$) is:

$$\Pi = \begin{cases} 
1 - \exp[-(s_m + s_f)] & \text{for } s_m + s_f > 0 \\
0 & \text{for } s_m + s_f < 0 
\end{cases}$$ \hspace{1cm} (8).

Whether a given mutation is eventually fixed, or is instead lost from the population, was determined by sampling from a Bernoulli distribution with parameter $\Pi$. For $0 < s_m + s_f << 1$, $\Pi \approx t_m + t_f$, which we often use in the analytical results, though never in the simulations.

Adaptive walks were simulated by sequentially fixing substitutions. Each substitution carries the population to a new set of distances from the optima ($z_m$ and $z_f$). The new angle between
selection vectors was calculated using the expressions provided above. We assumed that the phenotypic effect distribution of random mutations was unaffected by the evolving position of the population in phenotypic space \(i.e.,\) that the distribution of \(r_m\) and \(r_f\), and the similarity of mutant orientations in the two sexes, is insensitive to the fitness landscape).

In the model for the rate of adaptation (see above), we have the proportionality \(dz/dt \propto E[\Pi]\). Our analytical treatment makes use of a small selection coefficient approximation for the fixation probability, leading to \(dz/dt \propto E\{(t_m + t_f)/[1 – \exp(–N(t_m + t_f))]}\). In simulations we use the fixation probability in eq. (8). This leads to the proportionality, \(dz/dt \propto \Pr(s_{avg} > 0)E[\Pi_{ben}]\), where \(\Pr(s_{avg} > 0)\) is the fraction of random mutations that meet conditions for positive selection, and \(E[\Pi_{ben}]\) is the fixation probability among beneficial mutations. For each simulated data point, we generated 50,000 beneficial mutations and calculated \(\Pr(s_{avg} > 0) = 50,000/(\text{total mutations required to reach 50,000 beneficial})\), and \(E[\Pi_{ben}]\) as the average value of \(1 – \exp[–(s_m + s_f)]\), evaluated from the set of beneficial mutations.

References


Figure S1. Additional adaptive walk simulations. Blue and red curves (largely overlapping) refer to the strength of directional selection on females and males, respectively. The black line (± one standard deviation) shows \( \cos(\theta_{\text{sel}}) \) during the adaptive walk. Results show averages across 40 replicate adaptive walks, with parameters the same as in Fig. 3, unless otherwise stated.
Figure S2. Adaptive walk simulations, with differing fitness landscapes between the sexes: $\omega_m = 4/3$; $\omega_f = 2/3$. Blue and red curves refer to the strength of directional selection on females and males, respectively. The black line shows $\cos(\theta_{sel})$ during the adaptive walk. Parameters are the same as in Fig. 3, unless otherwise stated.